

# Projet court : Assignment des structures secondaire de protéines

## Introduction :

Une protéine est un arrangement tridimensionnelle d'acides aminés. Cet arrangement se fait sur quatre niveau structural. La structure primaire correspond à l'enchaînement des acides aminés de façon plane. La structure secondaire est un premier repliement qui existe grâce à des liaison hydrogène au sein d'une même chaîne polypeptidique et aux angles existants entre deux résidus. Elle comprend des hélices alpha et des feuillets bêta. La structure tertiaire se réfère tridimensionnelle où les repliements sont entraînés par des interactions hydrophobes et stabilisés par des liaisons hydrogènes et des ponts disulfures. La structure quaternaire est constituée de l'agrégation de deux ou plusieurs chaînes polypeptidiques qui forment alors une unité.

Les structures des protéines ont besoin d'être connues afin de pouvoir prédire leurs effets. S'intéresser à la structure tertiaire d'une protéine serait trop coûteuse en temps et en argent. Il vaut mieux donc s'intéresser à la structure secondaire. Celle-ci peut être prédite par différentes méthodes ou bien simplement être assigner à la protéine. L'assignation de la structure secondaire est possible grâce à plusieurs méthodes comme par exemple STRIDE, DEFINE ou DSSP. Nous nous intéresserons au dernier.

L'algorithme DSSP (dictionnaire de structure secondaire de protéine) est couramment utilisé. Il détermine la structure secondaire d'une protéine grâce aux liaisons hydrogène de celle-ci, selon le schéma initial proposé par Corey et Pauling en 1951. Le DSSP définit 8 types de structures secondaires différentes (hélice  $\alpha$ , coude, brin  $\beta$  en feuillet, pelote ... ). Il attribuera une lettre à chaque structure et c'est ainsi qu'il sera possible de les identifier pour les utilisateurs[1].

Ce projet avait pour but d'implémenter la méthode DSSP, en utilisant aussi les liaisons hydrogène d'une protéine pour en déduire sa structure secondaire et essayer de présenter les résultats de la même manière que le fait l'algorithme DSSP.

## Matériel et Méthode :

Le choix d'une protéine doit être fait et elle sera issue de la *Protein Data Bank* (PDB). Le fichier obtenu sera donc en format PDB. Ce PDB contient les informations concernant la protéine comme par exemple tous les atomes de cette protéine, leur coordonnées, le nom des acides aminés... Ce fichier PDB peut être utilisé sans aucun changement avec l'algorithme DSSP. Avec le nouvel algorithme créé, il faut préalablement traiter le fichier PDB avec le programme HBPLUS, qui doit être téléchargé.

HBPLUS est programme de calcul des liaison hydrogène développé par I.K MacDonald, D Naylor, D Jones et J.M Thornton à l'*University College London* [2]. Après l'obtention des autorisations, ce programme pourra analyser un fichier PDB et fournir un fichier hb2 qui comporte toutes les liaisons hydrogène détectées par le programme.

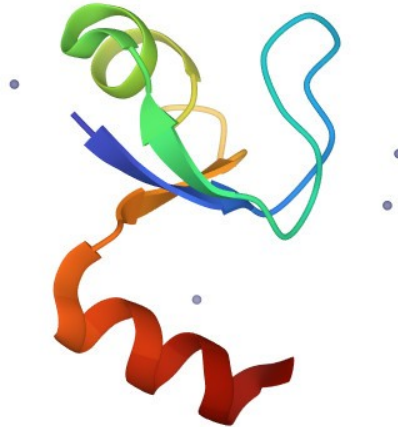
Le langage de programmation est python. L'interface utilisée est *Visual Studio Code*.

Pour pouvoir implémenter la méthode DSSP, l'article de Kabsch et Sander [3] a servi de base au code. En effet celui-ci définit clairement la façon de reconnaître les éléments composants les structures secondaire. Tout d'abord, le *n-turn* qui définit une liaison entre un résidu en position (i) et un autre en position (i+n). L'écart entre ces deux résidus est forcément de 3, 4 ou 5. Si l'on détecte des turn successifs, alors on peut dire que l'on a une hélice. Cette hélice est donc la première structure secondaire que l'on reconnaît dans l'article. Pour pouvoir détecter les feuillets bêta, il faut

d'abord définir les *bridges* puis les *ladders*, qui sont des *bridges* successifs. L'article expliquant toute la méthode pour reconnaître les différentes structures, il sert donc à notre méthode de programmation.

## Résultats :

La protéine choisie est 2HBB. C'est une protéine de 51 résidus, comportant deux hélices, trois *ladders* et un feuillet.



**Figure 1 : Structure 3D de la protéine 2HBB issue de PDB**

Le programme détecte bien les deux hélices situées entre les résidus 22 et 31 et entre les résidus 40 et 50. Il détecte également une partie des *ladders*.

21	-
22	hélice
23	hélice
24	hélice
25	hélice
26	hélice
27	hélice
28	hélice
29	hélice

39	-
40	hélice
41	hélice
42	hélice
43	hélice
44	hélice
45	hélice
46	hélice
47	hélice
48	hélice
49	hélice
50	hélice

**Figure 2 et 3 : Résultats obtenus pour les hélices avec le programme implémenté**

numéro	résidu	strusture
1		-
2		-
3		ladder
4		ladder
5		ladder
6		-
7		-

**Figure 4 : Résultat obtenu pour les ladders avec le programme implémenté**

## Discussion :

En comparant les résultats obtenus avec le programme crée et ceux obtenus avec DSSP, on peut remarquer l'imperfection du nouveau programme. En effet, celui-ci ne détecte pas tous les *ladders*. La non détection de certains *ladders* peut-être due au fait que le programme HBPLUS ne trouve pas de liaison hydrogène dans le fichier de base. Si les liaisons ne sont pas détectées, le programme ne peut pas les reconnaître comme *ladders* car elles n'existent pas pour lui.

En revanche, le programme a une bonne détection des hélices car il a plus de 90 % (22 résidus contre 24) de résidus formant une hélice en commun avec le fichier traité par DSSP (annexe 1 et 2).

Le programme crée mérite d'être retravaillé, amélioré et surtout optimisé afin que celui-ci soit plus simple d'utilisation et de compréhension du code.

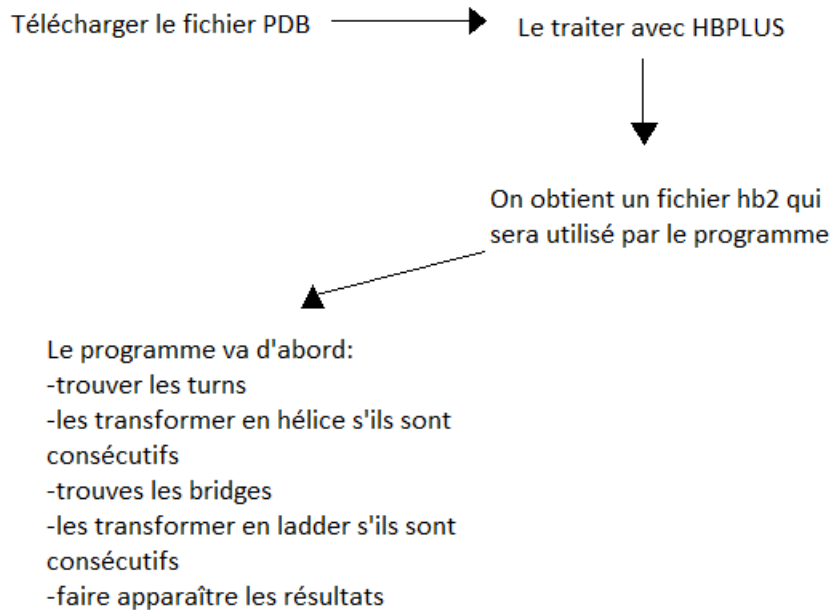
## Annexe :

#	RESIDUE	AA	STRUCTURE	BP1	BP2	i
1	1	A	M		0	0
2	2	A	K	E	+A	20 0A
3	3	A	V	E	-AB	19 38A
4	4	A	I	E	-AB	18 37A
5	5	A	F	E	- B	0 36A
6	6	A	L	S	S+	0 0
7	7	A	K	S	S-	0 0
8	8	A	D		-	0 0
9	9	A	V	B >>	-C	13 0B
10	10	A	K	T 34	S+	0 0

### **Annexe 1 : Résultat de DSSP pour les premiers ladders**

22	22	A	A	>	-	0	0
23	23	A	D	H	> S+	0	0
24	24	A	G	H	> S+	0	0
25	25	A	Y	H	>>S+	0	0
26	26	A	A	I	<>S+	0	0
27	27	A	N	I	<5S+	0	0
28	28	A	N	I	<5S+	0	0
29	29	A	F	I	X5S+	0	0
30	30	A	L	I	4XS+	0	0
31	31	A	F	T	>4<S+	0	0
32	32	A	K	T	345S+	0	0
33	33	A	Q	T	3<5S-	0	0
34	34	A	G	T	< 5S+	0	0
35	35	A	L	S	<S+	0	0

### **Annexe 2 : Résultat de DSSP pour la première hélice**



***Annexe 3 : Schéma reprenant le traitement de fichier avec le programme.***

Commentaire au sujet de la programmation en elle-même :

J'ai d'abord eu pas mal de problème pour extraire les liaison H du fichier. Je ne savais pas quel logiciel utiliser et avec Pymol je n'arrivais pas à récupérer un fichier pouvant ensuite être lu par python.

Une fois hbplus téléchargé, il a fallu un certain nombre de tentative avant de comprendre son fonctionnement.

Pour la partie codage, de nombreux schéma ont été nécessaire et l'article n'est pas toujours très clair. Mon programme est sûrement très simple et plutôt lourds à lire, mais c'est mon premier.

**Bibliographie :**

[1] [https://fr.wikipedia.org/wiki/Structure\\_secondaire#cite\\_note-2](https://fr.wikipedia.org/wiki/Structure_secondaire#cite_note-2)

[2] [http://www.csb.yale.edu/userguides/datamanip/hbplus/hbplus\\_descrip.html](http://www.csb.yale.edu/userguides/datamanip/hbplus/hbplus_descrip.html)

[3]

[https://git.pagelibre.org/richard/projet\\_court/raw/fc7e48c07f09fea182f10b9648acc4b43ba474ba/doc/kabsch1983.pdf](https://git.pagelibre.org/richard/projet_court/raw/fc7e48c07f09fea182f10b9648acc4b43ba474ba/doc/kabsch1983.pdf)